531632

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 April 2004 (29.04.2004)

PCT

(10) International Publication Number
WO 2004/036955 A1

(51) International Patent Classification[7]: H04S 7/00

(21) International Application Number:
PCT/KR2003/002149

(22) International Filing Date: 15 October 2003 (15.10.2003)

(25) Filing Language: Korean

(26) Publication Language: English

(30) Priority Data:
10-2002-0062962 15 October 2002 (15.10.2002) KR
10-2003-0071345 14 October 2003 (14.10.2003) KR

(71) Applicant *(for all designated States except US)*: **ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE** [KR/KR]; 161, Gajeong-dong, Yuseong-gu, Daejon 305-350 (KR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SEO, JEONG IL** [KR/KR]; #107-801 Sejong Apt., Jeonmin-dong, Yuseong-gu, Daejon 305-728 (KR). **JANG, DAE YOUNG** [KR/KR]; #101-1002 Hansol Apt., Songgang-dong, Yuseong-gu, Daejon 305-503 (KR). **KANG, KYEONG OK** [KR/KR]; #101-605, Samsungpuren Apt., Jeonmin-dong, Yuseong-gu, Daejon 305-727 (KR). **KIM, JIN WOONG** [KR/KR]; #305-1603, Expo Apt., Jeonmin-dong, Yuseong-gu, Daejon 305-761 (KR). **Ahn, Chieteuk** [KR/KR]; #208-603 Expo Apt., Jeonmin-dong, Yuseong-gu, Daejon 305-761 (KR).

(74) Agent: **SHINSUNG PATENT FIRM**; Haecheon Bldg., 741-40, Yeoksam 1-dong, Kangnam-gu, Seoul 135-924 (KR).

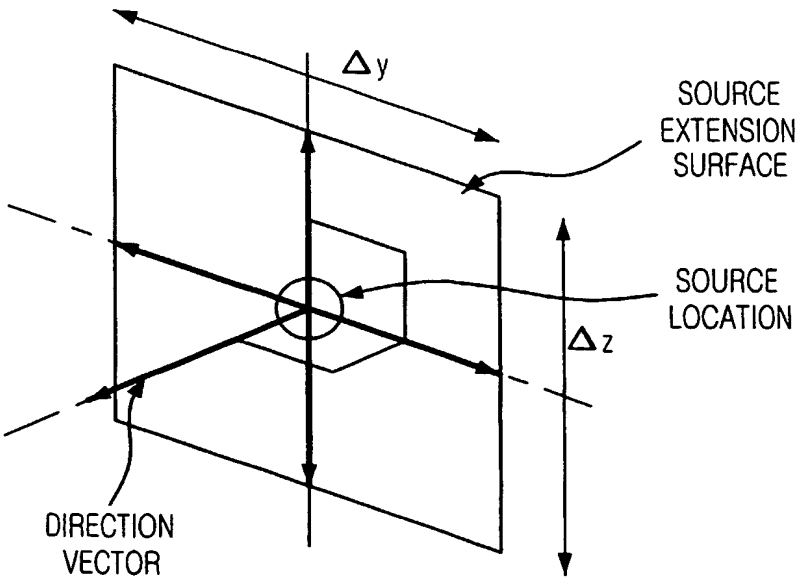
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *with international search report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: METHOD FOR GENERATING AND CONSUMING 3D AUDIO SCENE WITH EXTENDED SPATIALITY OF SOUND SOURCE




(57) Abstract: A method of generating and consuming 3D audio scene with extended spatiality of sound source describes the shape and size attributes of the sound source. The method includes the steps of: generating audio object; and generating 3D audio scene description information including attributes of the sound source of the audio object.

WO 2004/036955 A1

# METHOD FOR GENERATING AND CONSUMING 3-D AUDIO SCENE WITH EXTENDED SPATIALITY OF SOUND SOURCE

## Description

### Technical Field

The present invention relates to a method for generating and consuming a three-dimensional audio scene having sound source whose spatiality is extended; and, more particularly, to a method for generating and consuming a three-dimensional audio scene to extend the spatiality of sound source in a three-dimensional audio scene.

### Background Art

Generally, a content providing server encodes contents in a predetermined encoding method and transmits the encoded contents to content consuming terminals that consume the contents. The content consuming terminals decode the contents in a predetermined decoding method and output the transmitted contents.

Accordingly, the content providing server includes an encoding unit for encoding the contents and a transmission unit for transmitting the encoded contents. On the other hand, the content consuming terminals includes a reception unit for receiving the transmitted encoded contents, a decoding unit for decoding the encoded contents, and an output unit for outputting the decoded contents to users.

Many encoding/decoding methods of audio/video signals are known so far. Among them, an encoding/decoding method based on Moving Picture Experts Group 4 (MPEG-4) is widely used these days. MPEG-4 is a technical standard for data compression and restoration technology defined by the MPEG to transmit moving pictures at a low transmission rate.

According to MPEG-4, an object of an arbitrary shape

can be encoded and the content consuming terminals consume a scene composed of a plurality of objects. Therefore, MPEG-4 defines Audio Binary Format for Scene (Audio BIFS) with a scene description language for designating a sound object expression method and the characteristics thereof.

Meanwhile, along with the development in video, users want to consume contents of more lifelike sounds and video quality. In the MPEG-4 AudioBIFS, an AudioFX node and a DirectiveSound node are used to express spatiality of a three-dimensional audio scene. In these nodes, modeling of sound source is usually depended on point-source. Point-source can be described and embodied in a three-dimensional sound space easily.

Actual point-sources, however, tend to have a dimension more than two, rather than to be a point of literal meaning. More important thing here is that the shape of the sound source can be recognized by human beings, which is disclosed by J. Baluert, "Spatial Hearing," the MIT Press, Cambridge Mass, 1996.

For example, a sound of waves dashing against the coastline stretched in a straight line can be recognized as a linear sound source instead of a point sound source. To improve the sense of the real of the three-dimensional audio scene by using the AudioBIFS, the size and shape of the sound source should be expressed. Otherwise, the sense of the real of a sound object in the three-dimensional audio scene would be damaged seriously.

That is, the spatiality of a sound source could be described to endow a three-dimensional audio scene with a sound source which is of more than one-dimensional.

## Disclosure of Invention

It is, therefore, an object of the present invention to provide a method for generating and consuming a three-

dimensional audio scene having a sound source whose spatiality is extended by adding sound source characteristics information having information on extending the spatiality of the sound source to three-dimensional audio scene description information.

The other objects and advantages of the present invention can be easily recognized by those of ordinary skill in the art from the drawings, detailed description and claims of the present specification.

In accordance with one aspect of the present invention, there is provided a method for generating a three-dimensional audio scene with a sound source whose spatiality is extended, including the steps of: a) generating a sound object; and b) generating three-dimensional audio scene description information including sound source characteristics information for the sound object, wherein the sound source characteristics information includes spatiality extension information of the sound source which is information on the size and shape of the sound source expressed in a three-dimensional space.

In accordance with one aspect of the present invention, there is provided a method for consuming a three-dimensional audio scene with a sound source whose spatiality is extended, including the steps of: a) receiving a sound object and three-dimensional audio scene description information including sound source characteristics information for the sound object; and b) outputting the sound object based on the three-dimensional audio scene description information, wherein the sound source characteristics information includes spatiality extension information which is information on the size and shape of a sound source expressed in a three-dimensional space.

## Brief Description of Drawings

The above and other objects and features of the present invention will become apparent from the following description of the preferred embodiments given in conjunction with the accompanying drawings, in which:

Fig. 1 is a diagram illustrating various shapes of sound sources;

Fig. 2 is a diagram describing a method for expressing spatial sound source by grouping successive point sound sources;

Fig. 3 shows an example where spatiality extension information is added to a "DirectiveSound" node of AudioBIFS in accordance with the present invention;

Fig. 4 is a diagram illustrating how a sound source is extended in accordance with the present invention; and

Fig. 5 is a diagram depicting the distributions of point sound sources based on the shapes of various sound sources in accordance with the present invention.

## Best Mode for Carrying Out the Invention

Other objects and aspects of the invention will become apparent from the following description of the embodiments with reference to the accompanying drawings, which is set forth hereinafter.

Following description exemplifies only the principles of the present invention. Even if they are not described or illustrated clearly in the present specification, one of ordinary skill in the art can embody the principles of the present invention and invent various apparatuses within the concept and scope of the present invention.

The use of the conditional terms and embodiments presented in the present specification are intended only to make the concept of the present invention understood, and

they are not limited to the embodiments and conditions mentioned in the specification.

In addition, all the detailed description on the principles, viewpoints and embodiments and particular embodiments of the present invention should be understood to include structural and functional equivalents to them. The equivalents include not only currently known equivalents but also those to be developed in future, that is, all devices invented to perform the same function, regardless of their structures.

For example, block diagrams of the present invention should be understood to show a conceptual viewpoint of an exemplary circuit that embodies the principles of the present invention. Similarly, all the flowcharts, state conversion diagrams, pseudo codes and the like can be expressed substantially in a computer-readable media, and whether or not a computer or a processor is described distinctively, they should be understood to express various processes operated by a computer or a processor.

Functions of various devices illustrated in the drawings including a functional block expressed as a processor or a similar concept can be provided not only by using hardware dedicated to the functions, but also by using hardware capable of running proper software for the functions. When a function is provided by a processor, the function may be provided by a single dedicated processor, single shared processor, or a plurality of individual processors, part of which can be shared.

The apparent use of a term, 'processor', 'control' or similar concept, should not be understood to exclusively refer to a piece of hardware capable of running software, but should be understood to include a digital signal processor (DSP), hardware, and ROM, RAM and non-volatile memory for storing software, implicatively. Other known and commonly used hardware may be included therein, too.

In the claims of the present specification, an element expressed as a means for performing a function described in the detailed description is intended to include all methods for performing the function including all formats of software, such as combinations of circuits for performing the intended function, firmware/microcode and the like. To perform the intended function, the element is cooperated with a proper circuit for performing the software. The present invention defined by claims includes diverse means for performing particular functions, and the means are connected with each other in a method requested in the claims. Therefore, any means that can provide the function should be understood to be an equivalent to what is figured out from the present specification.

Other objects and aspects of the invention will become apparent from the following description of the embodiments with reference to the accompanying drawings, which is set forth hereinafter. The same reference numeral is given to the same element, although the element appears in different drawings. In addition, if further detailed description on the related prior arts is determined to blur the point of the present invention, the description is omitted. Hereafter, preferred embodiments of the present invention will be described in detail.

Fig. 1 is a diagram illustrating various shapes of sound sources. Referring to Fig. 1, a sound source can be a point, a line, a surface and space having a volume. Since sound source has an arbitrary shape and size, it is very complicated to describe the sound source. However, if the shape of the sound source to be modeled is controlled, the sound source can be described less complicatedly.

In the present invention, it is assumed that point sound sources are distributed uniformly in the dimension of a virtual sound source in order to model sound sources of various shapes and sizes. As a result, the sound sources

of various shapes and sizes can be expressed as continuous arrays of point sound sources. Here, the location of each point sound source in a virtual object can be calculated using a vector location of a sound source which is defined in a three-dimensional scene.

When a spatial sound source is modeled with a plurality of point sound sources, the spatial sound source should be described using a node defined in AudioBIFS. When the node defined in AudioBIFS, which will be referred to as an AudioBIFS node, is used, any effect can be included in the three-dimensional scene. Therefore, an effect corresponding to the spatial sound source can be programmed through the AudioBIFS node and inserted to the three-dimensional scene.

However, this requires very complicated Digital Signal Processing (DSP) algorithm and it is very troublesome to control the dimension of the spatial sound source.

Also, the point sound sources distributed in a limited dimension of an object are grouped using the AudioBIFS, and the spatial location and direction of the sound sources can be changed by changing the sound source group. First of all, the characteristics of the point sound sources are described using a plurality of "DirectiveSound" node. The locations of the point sound sources are calculated to be distributed on the surface of the object uniformly.

Subsequently, the point sound sources are located with a spatial distance that can eliminate spatial aliasing, which is disclosed by A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," J. Aoust. Soc. Am., Vol. 93, No. 5 on pages from 2764 to 2778, May, 1993. The spatial sound source can be vectorized by using a group node and grouping the point sound sources.

Fig. 2 is a diagram describing a method for expressing spatial sound source by grouping successive point sound

sources. In the drawing, a virtual successive linear sound source is modeled by using three point sound sources which are distributed uniformly along the axis of the linear sound source.

The locations of the point sound sources are determined to be $(x_0-dx, y_0-dy, z_0-dz)$, $(x_0, y_0, z_0)$, and $(x_0+dx, y_0+dy, z_0+dz)$ according to the concept of the virtual sound source. Here, dx, dy and dz can be calculated from a vector between a listener and the location of the sound source and the angle between the direction vectors of the sound source, the vector and the angle which are defined in an angle field and a direction field.

Fig. 2 describes a spatial sound source by using a plurality of point sound sources. AudioBIFS appears it can support the description of a particular scene. However, this method requires too much unnecessary sound object definition. This is because many objects should be defined to model one single object.

When it is told that the genuine object of hybrid description of Moving Picture Experts Group 4 (MPEG-4)is more object-oriented representations, it is desirable to combine the point sound sources, which are used for model one spatial sound source, and reproduce one single object.

In accordance with the present invention, a new field is added to a "DirectiveSound" node of the AudioBIFS to describe the shape and size attributes of a sound source. Fig. 3 shows an example where spatiality extension information is added to a "DirectiveSound" node of AudioBIFS in accordance with the present invention.

Referring to Fig. 3, a new rendering design corresponding to a value of a "SourceDimensions" field is applied to the "DirectiveSound" node. The "SourceDimensions" field also includes shape information of the sound source. If the value of the "SourceDimensions"

field is "0,0,0", the sound source becomes one point, no additional technology for extending the sound source is applied to the "DirectiveSound" node. If the value of the "SourceDimensions" field is a value other than "0,0,0", the dimension of the sound source is extended virtually.

The location and direction of the sound source are defined in a location field and a direction field, respectively, in the "DirectiveSound" node. The dimension of the sound source is extended in vertical to a vector defined in the direction field based on the value of the "SourceDimensions" field.

The "location" field defines the geometrical center of the extended sound source, whereas the "SourceDimensions" field defines the three-dimensional size of the sound source. In short, the size of the sound source extended spatially is determined according to the values of $\Delta x$, $\Delta y$ and $\Delta z$.

Fig. 4 is a diagram illustrating how a sound source is extended in accordance with the present invention. As illustrated in the drawing, the value of the "SourceDimensions" field is (0, $\Delta y$, $\Delta z$), $\Delta y$ and $\Delta z$ being not zero ($\Delta y \neq 0$, $\Delta z \neq 0$). This indicates a surface sound source having an area of $\Delta y \times \Delta z$.

The illustrated sound source is extended in a direction vertical to a vector defined in the "direction" field based on the values of the "SourceDimensions" field, i.e., (0, $\Delta y$, $\Delta z$), and thereby forming a surface sound source. As shown in the above, when the dimension and location of a sound source is defined, the point sound sources are located on the surfaces of the extended sound source. In the present invention, the locations of the point sound sources are calculated to be distributed on the surfaces of the extended sound source uniformly.

Figs. 5A to 5C are diagrams depicting the

distributions of point sound sources based on the shapes of various sound sources in accordance with the present invention. The dimension and distance of a sound source are free variables. So, the size of the sound source that can be recognized by a user can be formed freely.

For example, multi-track audio signals that are recorded by using an array of microphones can be expressed by extending point sound sources linearly as shown in Fig. 5A. In this case, the value of the "SourceDimensions" field is $(0, 0, \Delta z)$.

Also, different sound signals can be expressed as an extension of a point sound source to generate a spread sound source. Figs. 5B and 5C show a surface sound source expressed through the spread of the point sound source and a spatial sound source having a volume. In case of Fig. 5B, the value of the "SourceDimensions" field is $(0, \Delta y, \Delta z)$ and, in case of Fig. 5C, the value of the "SourceDimensions" field is $(\Delta x, \Delta y, \Delta z)$.

As the dimension of a spatial sound source is defined as described in the above, the number of the point sound sources (i.e., the number of input audio channels) determines the density of the point sound sources in the extended sound source.

If an "AudioSource" node is defined in a "source" field, the value of a "numChan" field may indicate the number of used point sound sources. The directivity defined in "angle," "directivity" and "frequency" fields of the "DirectiveSound" node can be applied to all point sound sources included in the extended sound source uniformly.

The apparatus and method of the present invention can produce more effective three-dimensional sounds by extending the spatiality of sound sources of contents.

While the present invention has been described with respect to certain preferred embodiments, it will be

apparent to those skilled in the art that various changes and modifications may be made without departing from the scope of the invention as defined in the following claims.

What is claimed is:

1. A method for generating a three-dimensional audio scene with a sound source whose spatiality is extended, comprising the steps of:

a) generating a sound object; and

b) generating three-dimensional audio scene description information including sound source characteristics information for the sound object,

wherein the sound source characteristics information includes spatiality extension information of the sound source which is information on the size and shape of the sound source expressed in a three-dimensional space.

2. The method as recited in claim 1, wherein the spatiality extension information of the sound source includes sound source dimension information that is expressed as an x component, y component and z component of a three-dimensional rectangular coordinates.

3. The method as recited in claim 2, wherein the spatiality extension information of the sound source further includes geometrical center location information of the sound source dimension information.

4. The method as recited in claim 2, wherein the spatiality extension information of the sound source further includes direction information of the sound source and describes a three-dimensional audio scene by extending the spatiality of the sound source in a direction vertical to the direction of the sound source.

5. A method for consuming a three-dimensional audio scene with a sound source whose spatiality is extended, comprising the steps of:

a) receiving a sound object and three-dimensional audio scene description information including sound source characteristics information for the sound object; and

b) outputting the sound object based on the three-dimensional audio scene description information,

wherein the sound source characteristics information includes spatiality extension information which is information on the size and shape of the sound source expressed in a three-dimensional space.

6. The method as recited in claim 5, wherein spatiality extension information of the sound source includes sound source dimension information that is expressed as an x component, y component and z component of a three-dimensional rectangular coordinates.

7. The method as recited in claim 6, wherein the spatiality extension information of the sound source further includes geometrical center location information of the sound source dimension information.

8. The method as recited in claim 6, wherein the spatiality extension information of the sound source further includes direction information of the sound source and describes a three-dimensional audio scene by extending the spatiality of the sound source in a direction vertical to the direction of the sound source.

9. A three-dimensional audio scene data stream with a sound source whose spatiality is extended, comprising:

a sound object; and

three-dimensional audio scene description information including sound source characteristics information for the sound object data,

wherein the sound source characteristics information

includes spatiality extension information which is information on the size and shape of the sound source expressed in a three-dimensional space.

10. The data stream as recited in claim 9, wherein the spatiality extension information of the sound source includes sound source dimension information that is expressed as an x component, y component and z component of a three-dimensional rectangular coordinates.

11. The data stream as recited in claim 9, wherein the spatiality extension information of the sound source further includes geometrical center location information of the sound source dimension information.

12. The data stream as recited in claim 9, wherein the spatiality extension information of the sound source further includes direction information of the sound source and describes a three-dimensional audio scene by extending the spatiality of the sound source in a direction vertical to the direction of the sound source.

FIG. 1A

FIG. 1B

FIG. 1C

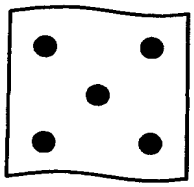


FIG. 1D

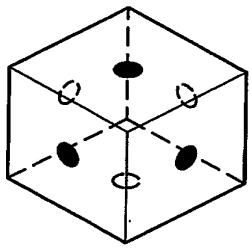


FIG. 2

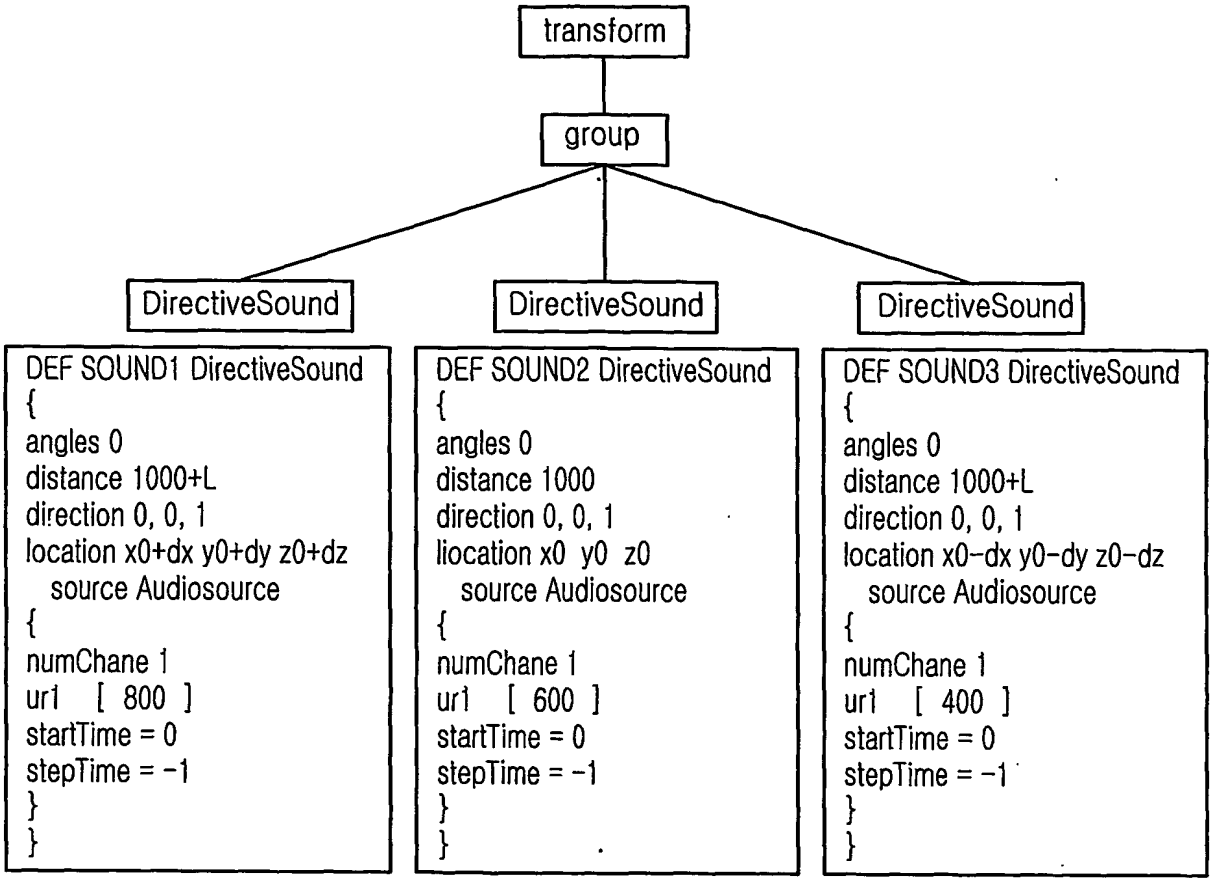

transform
group
DirectiveSound
DirectiveSound
DirectiveSound
DEF SOUND1 DirectiveSound
{
angles 0
distance 1000+L
direction 0, 0, 1
location x0+dx y0+dy z0+dz
source Audiosource
{
numChane 1
url [ 800 ]
startTime = 0
stepTime = -1
}
}
DEF SOUND2 DirectiveSound
{
angles 0
distance 1000
direction 0, 0, 1
liocation x0 y0 z0
source Audiosource
{
numChane 1
url [ 600 ]
startTime = 0
stepTime = -1
}
}
DEF SOUND3 DirectiveSound
{
angles 0
distance 1000+L
direction 0, 0, 1
location x0-dx y0-dy z0-dz
source Audiosource
{
numChane 1
url [ 400 ]
startTime = 0
stepTime = -1
}
}

FIG. 3

DirectiveSound {

| | | | |
|---|---|---|---|
| Field | MFFloat | angles | 0 |
| Field | MFFloat | directivity | 1 |
| Field | MFFloat | frequency | [] |
| Field | SFFloat | speedOfSound | 340 |
| Field | SFFloat | distance | 100 |
| Field | SFBool | useAirabs | FALSE |
| exposedField | SFVec3f | direction | 0, 0, 1 |
| exposedField | SFFloat | intensity | 1 |
| exposedField | SFVec3f | location | 0, 0, 0 |
| exposedField | SFVec3f | sourceDimensions | 0, 0, 0 |
| exposedField | SFNode | PerceptualParameters | NULL |
| exposedField | SFBool | RoomEffect | FALSE |
| exposedField | SFBool | Spatialize | TRUE |

FIG. 4

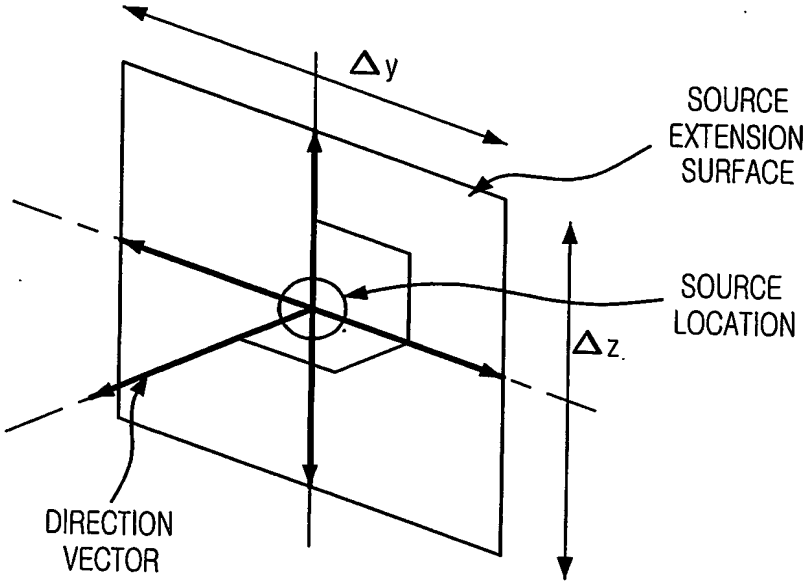



FIG. 5A

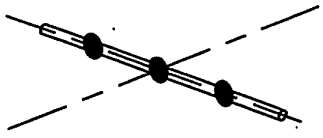


FIG. 5B

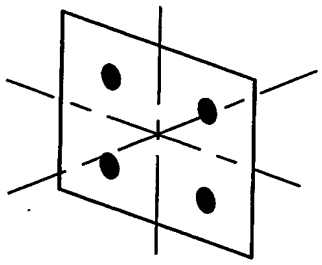
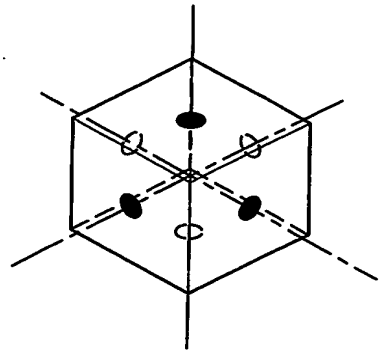


FIG. 5C



# INTERNATIONAL SEARCH REPORT

rnational application No.
PCT/KR2003/002149

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC7 H04S 7/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04S, H04R, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

KR, JP: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | US 20010037386A1 (Susumu Takatsuka et.al.) 01 NOVEMBER. 2001<br>See the whole document | 1 |
| A | US 20010043738A1 (Harpreet Singh Sawhney et.al.) 22 NOVEMBER. 2001<br>See the whole document | 1 |
| A | KR 20010028744A (Mahn-jin Han et.al.) 11 OCTOBER. 2001<br>See the whole document | 1 |

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 20 JANUARY 2004 (20.01.2004) | 20 JANUARY 2004 (20.01.2004) |


Name and mailing address of the ISA/KR
Korean Intellectual Property Office
920 Dunsan-dong, Seo-gu, Daejeon 302-701,
Republic of Korea
Facsimile No. 82-42-472-7140

Authorized officer
KIM, Seung Jo
Telephone No. 0421)481-5675

Information on patent family members

International application No.
PCT/KR2003/002149

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| US 20010037386A1 | 01-11-01 | None | |
| US 20010043738A1 | 22-11-01 | None | |
| KR 20010028744A | 11-10-01 | None | |